Human-robot co-carrying using visual and force sensing

Article (Accepted Version)

Human-Robot Co-Carrying Using Visual and Force Sensing

Xinbo Yu, *Member, IEEE,* Wei He, *Senior Member, IEEE,* Qing Li, Yanan Li, *Member, IEEE*, Bin Li

***Abstract*—In this paper, we propose a hybrid framework using visual and force sensing for human-robot co-carrying tasks. Visual sensing is utilized to obtain human motion and an observer is designed for estimating control input of human, which generates robot's desired motion towards human's intended motion. An adaptive impedance-based control strategy is proposed for trajectory tracking with neural networks (NNs) used to compensate for uncertainties in robot's dynamics. Motion synchronization is achieved and this approach yields a stable and efficient interaction behavior between human and robot, decreases human control effort and avoids interference to human during the interaction. The proposed framework is validated by a co-carrying task in simulations and experiments.**

***Index Terms*—Human-robot collaboration, Motion synchronization, Observer, Neural networks, Visual and force sensing.**

I. Introduction

Physical interaction of human and robot (pHRI) in shared environments and joint tasks poses many challenges [1]. There exist extensive applications of pHRI found in service and industrial areas including assembly, rehabilitation [2] and so on [3]. Co-carrying tasks, which rely on complementary advantages of human and robot, cannot be accomplished individually by a single human or robot. Coupled relationships between robot, transported object and human bring difficulties in analyzing behaviours of both human and robot. In this paper, the focus of interest is using vision and force sensing together to enable human and robot collaboratively to perform a co-carrying task. For this purpose, robot should have the following abilities:

1) measure *human motion* and estimate *human motion intention* for achieving motion synchronization;
2) measure external force on robot gripper and regulate interaction force for achieving safe interaction;
3) carry out human-in-the-loop control strategy considering system uncertainties.

This work was supported in part by the National Natural Science Foundation of China under Grant 61933001, 61621063, in part by Beijing Top Discipline for Artificial Intelligent Science and Engineering, University Science and Technology Beijing, and in part by the Fundamental Research Funds for the China Central Universities of USTB under Grant FRF-TP-19-001C2.

X. Yu is working with the Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing 100083, China. W. He and B. Li are with the School of Automation and Electrical Engineering, Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing 100083, China. Q. Li is working with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China. Y. Li is with the Department of Engineering and Design, University of Sussex, Brighton, BN1 9RH, UK.

The corresponding author is W. He, Email: weihe@ieee.org.

How to understand human sensorimotor behavior is a key for robot to achieve compliant interaction in pHRI. Robot can take a "*follower*" role in pHRI without knowing human motion intention. However, passive control of robot may disturb human behaviors, affect human trust or bring human more burden in co-transporting tasks. In [4], online neural networks are employed to estimate human motion intention which is defined as the human's desired trajectory, and less human efforts are required with the proposed method. In [5], human motion intention is estimated by observing robot control effort without force sensors. A switching control scheme is developed that changes between impedance control and interaction control. In [6], human motion intention is deemed as the current *human motion*, without considering the future motion estimation. Human muscle activity measurements encode the information about human motion, and provide robot with online feedback information. Based on this idea, human motion intention also can be defined as continuously time-varying force or torque [7]. In our paper, we regard human motion intention as moving target position which leads to a continuous trajectory. A related work can be found in [8], which can provide a desired trajectory based on the interaction force in pHRI without constant human guidance and results in reduction in human control effort. Most of these studies in the field of motion intention estimation have only focused on direct interaction rather than indirect pHRI, i.e. through an object in co-carrying tasks. Complicated coupled relationships between human, object and robot bring more difficulties in estimating human motion intention. Therefore one purpose of our work is to estimate human motion intention in co-carrying tasks with indirect interaction.

Controller design in pHRI has received considerable attentions in recent years [9]–[16]. A remarkable issue is that accurate robot's dynamics are extremely difficult to obtain from the engineering point of view [17]. However, it is critical to acquire sufficient information about robot's dynamics for achieving precise torque control [18]–[21]. In [22], an adaptive impedance control of dual-arm robots is proposed where neural networks (NNs) are utilized to compensate for uncertain dynamics. In [23], human-like adaptive controller is proposed for compensating for disturbance and dynamics without force sensing, and it is derived from minimizing control effort and error. In conclusion, uncertainty compensation is an important component in controller design [24]–[29], and also plays a key role in pHRI [30].

Joint tasks have been extensively studied in the field of pHRI [31]. In [32], authors design a control strategy which allows a humanoid robot to perform a complex co-carrying

task with human, and robot can guess human motion to proactively participate to the task. Human stiffness estimation is also an increasingly important area in co-carrying tasks. In [33], weight least-squares estimation is employed to estimate virtual stiffness, which is included in the complete set of task-parameterized Gaussian mixture model. This model can be used for impedance-based behaviors transfer. A related work can be found in [34], where authors proposed an approach considering probabilistic stiffness estimation, and encoding robot behavior in the task involving physical contact with the human. In [35], unknown grasp pose of human is identified, and online estimation of relative kinematics is derived by least square method. In the subsequent works, the authors consider that the object dynamics are unknown in co-transporting tasks, so they propose an identification strategy of object dynamics in the condition that inputs satisfy persistence of excitation (PE) [36]. However, most of research relies on force information in human control input. In [37], a hybrid controller combining visual servoing and impedance controller is considered in the task of joint carrying and a "ball-on-plate" system is employed to validate the effectiveness of this controller. In [38], an observer is designed to estimate control input of human, and motion synchronization in a direct pHRI scenario is achieved without requiring force sensory information at the interaction point. Inspired by the aforementioned works [37] and [38], visual servoing and observer are employed in our controller design. Visual servoing is utilized to obtain human motion and observer is used in estimating control input of human. A hybrid framework including visual and force sensing is proposed for human-robot co-carrying tasks. In [39], a companion robot is designed to switch between the visual servoing and force servoing modes, different from our framework using visual and force sensing to estimate the human motion intention. In [40], [41], robots learn the teaching-learning-collaboration model and predict human motion through learning by demonstrations or using historical data for training, while our method requires neither of them. Based on previous discussions, we highlight our contributions as follows:

1) A hybrid framework using visual and force sensing is proposed for human-robot co-carrying tasks, enabling the robot to proactively follow its human partner and reduce their control effort;
2) A force observer is designed to estimate human force without using the force sensor, and human motion intention is obtained by minimizing the estimated force.

The proposed framework includes both visual and force sensing, and a controller combining visual servoing control and impedance control is designed, so we call it "hybrid".

The rest of the paper is presented as follows: Section II presents robot and object dynamics, and control inputs of human and robot are analyzed; Section III introduces the proposed method; Section IV and Section V evaluate its performance by simulations and experiments; Section VI and Section VII conclude this work and discuss future works.

II. Problem Formulation

A. System description

We consider a co-carrying task where human and robot transport a rigid object as depicted in Fig. 1. All vectors and matrices are defined in the fixed coordinate frame of robot $\{R\}$, of which the origin is at the robot's mass center.

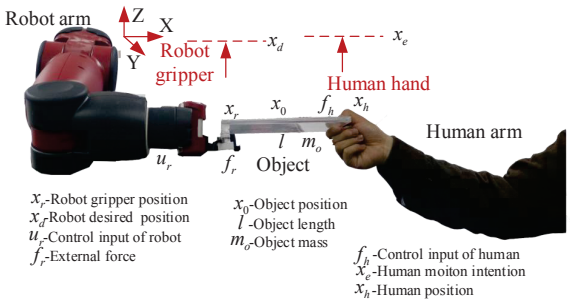


Fig. 1: Human-robot co-carrying task.

The dynamics of an n link robot is given by:

$$M_r(x_r)\ddot{x}_r + C_r(x_r, \dot{x}_r)\dot{x}_r + G_r(x_r) = u_r - f_r \quad (1)$$

where $M_r(x_r) \in \mathbb{R}^{n\times n}, C_r(x_r, \dot{x}_r) \in \mathbb{R}^{n\times n}, G_r(x_r) \in \mathbb{R}^n$ denote mass matrix, Coriolis and centripetal matrix, gravity vector in robot's dynamics, respectively; $x_r, \dot{x}_r, \ddot{x}_r \in \mathbb{R}^n$ denote position, velocity and acceleration vectors of robot gripper, respectively; $f_r \in \mathbb{R}^n$ denotes external force on the robot measured by a force sensor or calculated based on torque sensors, and $u_r \in \mathbb{R}^{n\times n}$ denotes the control input of robot. Similarly, the dynamics of an n dimension transported object can be described as follows:

$$M_o(x_o)\ddot{x}_o + C_o(x_o, \dot{x}_o)\dot{x}_o + G_o(x_o) = f_h + f_r \quad (2)$$

where $M_o(x_o) \in \mathbb{R}^{n\times n}$, $C_o(x_o, \dot{x}_o) \in \mathbb{R}^{n\times n}$ and $G_o(x_o) \in \mathbb{R}^n$ denote mass matrix, Coriolis and centripetal matrix and gravity vector in object's dynamics, respectively; and x_o, $\dot{x}_o$ and $\ddot{x}_o \in \mathbb{R}^n$ denote position, velocity and acceleration vectors of object's mass center; $f_h \in \mathbb{R}^n$ denotes the external force at the grasp point onto the transported object.

B. Control input

1) Control input of human: In this work, *control input of human* is defined as f_h in object dynamics model (2). We describe it as a simplified stiffness model as follows:

$$f_h = -K_h(x_h - x_e) \quad (3)$$

where $K_h \in \mathbb{R}^{n\times n}$ denotes the stiffness matrix of human arm, $x_h \in \mathbb{R}^n$ denotes human position, i.e., position vector of human hand which contacts with the transported object, $x_e \in \mathbb{R}^n$ denotes human motion intention.

We can see that if robot knows human motion intention x_e in real time and moves towards the desired motion x_d according to x_e, the co-carrying task can be performed successfully and *actively*. According to (3), if f_h, x_h and K_h are obtained, x_e can be calculated. Before we design a tracking control algorithm for robot, some issues should be addressed about how to acquire f_h, x_h and K_h:

1) In a scenario where it is infeasible to directly measure human force, how we can obtain f_h should be considered;
2) There are differences between x_h and x_r when human and robot collaborate to transport the object, so how we obtain x_h should be addressed;
3) Without knowledge of human stiffness K_h, an effective estimation method to obtain x_e should be proposed.

2) Control input of robot: Different from the definition of *control input of human*, *control input of robot* is u_r rather than the external force f_r on the object. In view of control objective, the control input of robot is designed as follows:

$$u_r = u_{rf} + u_{rb} + u_{ri} \tag{4}$$

where u_{rf} denotes the feedforward input for compensating for robot's dynamics, u_{rb} denotes the feedback input for tracking the desired trajectory x_d, and u_{ri} denotes the input for compensating for the interaction force. We can design corresponding inputs as follows:

$$\begin{aligned} u_{rf} &= M_r(x_r)\ddot{x}_d + C_r(x_r, \dot{x}_r)\dot{x}_d + G_r(x_r) \\ u_{rb} &= -K_P(x_r - x_d) - K_D(\dot{x}_r - \dot{x}_d) + K_Q \text{sgn}(\dot{x}_r - \dot{x}_d) \\ u_{ri} &= f_r + ((\dot{x}_r - \dot{x}_d)^T)^+(\dot{x}_h - \dot{x}_d)\hat{f}_h \end{aligned} \tag{5}$$

where K_P denotes the proportional gain matrix, and K_D denotes the differential gain matrix, which can be interpreted as stiffness and damping matrices in impedance control. K_Q will be explained subsequently in stability analysis. $((\dot{x}_r - \dot{x}_d)^T)^+$ denotes the Moore-Penrose inverse of $(\dot{x}_r - \dot{x}_d)^T$, and sgn$(\cdot)$ returns a vector with the signs of the corresponding elements of the vector $(\cdot)$. To address uncertainties in robot's dynamics, neural networks (NNs) are employed to compensate for them. And we design the weight adaptive law as follows:

$$\dot{\hat{\theta}}_i = -\Gamma_i[S_i(Z_i)(\dot{x}_{ri} - \dot{x}_{di}) + \sigma_i\hat{\theta}_i], i = 1, 2, ..., n \tag{6}$$

where $\hat{\theta}_i$ denotes weight estimates of NN, $\Gamma_i = \Gamma_i^T$ denotes a positive definite matrix and σ_i denotes small positive constants. Z_i=[x_r^T, $\dot{x}_r^T$, $\dot{x}_d^T$, $\ddot{x}_d^T$] denotes the input of NN, $S_i(Z_i)$ denotes basis functions, and $\hat{\theta}^T S(Z)$ is used to estimate $\theta^{*T} S(Z)$ as below

$$\theta^{*T} S(Z) = (M_r(x_r)\ddot{x}_d + C_r(x_r, \dot{x}_r)\dot{x}_d + G_r(x_r)) - \varepsilon(Z) \tag{7}$$

where θ_i^* denotes actual weight of NN, and the estimation error vector $\varepsilon(Z)$ stays in bounds over the compact set Ω_ε, $\forall Z \in \Omega_\varepsilon$, $||\varepsilon(Z)|| < \bar{\varepsilon}$, with $\bar{\varepsilon}$ as a positive constant. Then we develop adaptive NN as follows:

$$\begin{aligned} u_{rNN} = \hat{\theta}^T S(Z) - K_P(x_r - x_d) - K_D(\dot{x}_r - \dot{x}_d) + f_r \\ + K_Q \text{sgn}(\dot{x}_r - \dot{x}_d) + ((\dot{x}_r - \dot{x}_d)^T)^+(\dot{x}_h - \dot{x}_d)\hat{f}_h \end{aligned} \tag{8}$$

Substituting (8) to (1), we can obtain the closed-loop error dynamics:

$$\begin{aligned} M_r(x_r)\ddot{e} + (C_r(x_r, \dot{x}_r) + K_D)\dot{e} + K_P e = -\theta^{*T} S(Z) - \\ \varepsilon(Z) + \hat{\theta}^T S(Z) + K_Q \text{sgn}(e) + (\dot{e}^T)^+(\dot{x}_h - \dot{x}_d)\hat{f}_h \end{aligned} \tag{9}$$

where $e \in \mathbb{R}^n$ denotes the tracking error between x_r and x_d. In (8), if we set x_d as x_e, which means that robot is aware of human motion intention, robot can conduct *synchronous* and *active* collaboration with human partner in the co-transporting task. If we want to update x_d relative to x_e, we need to obtain human motion x_h by visual sensing according to analysis in Section III.

III. Sensing, observer design and human motion intention estimation

A. Visual sensing and force sensing

1) Visual sensing: We use a visual sensor to obtain human motion x_h in our work. x_h obtained by the visual sensor should be transformed into robot coordinate $\{R\}$ for further controller design. Calibration systems are employed to obtain the relationship between robot coordinate $\{R\}$ and camera coordinate $\{K\}$, so a calibration board is fixed on the robot's end-effector. The transformation matrix ${}^R_K T$ denotes the camera coordinate with respect to robot reference coordinate, matrix ${}^R_E T$ denotes the robot's end-effector coordinate with respect to robot reference coordinate, matrix ${}^E_C T$ denotes the calibration board coordinate with respect to robot's end-effector coordinate, and matrix ${}^C_K T$ denotes camera coordinate with respect to calibration board coordinate. The transformation matrix ${}^R_K T$ can be obtained as follows:

$${}^R_K T = {}^R_E T \cdot {}^E_C T \cdot {}^C_K T = {}^R_E T \cdot {}^E_C T \cdot ({}^K_C T)^{-1} \tag{10}$$

which can be rewritten as follows:

$${}^E_C T = ({}^R_E T)^{-1} \cdot {}^R_K T \cdot {}^K_C T \tag{11}$$

Considering that ${}^E_C T$ has no change when robot is in different poses, transformed coordinate matrices in two different poses are given as follows:

$$({}^R_E T_1)^{-1} \cdot {}^R_K T \cdot {}^K_C T_1 = ({}^R_E T_2)^{-1} \cdot {}^R_K T \cdot {}^K_C T_2 \tag{12}$$

where *1* and *2* denote two different poses. Then we rewrite (12) as follows:

$${}^R_E T_2 \cdot ({}^R_E T_1)^{-1} \cdot {}^R_K T = {}^R_K T \cdot {}^K_C T_2 \cdot ({}^K_C T_1)^{-1} \tag{13}$$

where we define that

$$X = {}^R_K T, A = {}^R_E T_2 \cdot ({}^R_E T_1)^{-1}, B = {}^K_C T_2 \cdot ({}^K_C T_1)^{-1} \tag{14}$$

where we can utilize a numerical method to solve X in $AX = XB$. Solving X is not the focus. In particular we use the well-known method in [42] to obtain ${}^R_K T$. By fixing the calibration board at the human hand position on the object, we can transform human motion from camera coordinate $\{K\}$ to robot reference coordinate $\{R\}$:

$${}^R x_h = {}^R_K T \cdot {}^K x_h \tag{15}$$

where ${}^K x_h$ denotes the human motion in camera coordinate $\{K\}$, and ${}^R x_h$ denotes the human motion in robot reference coordinate $\{R\}$, respectively. We notice that in practical applications, moving average filter should be utilized because calibration board may not be recognized during a task. Filters can provide predicted data $\hat{x}_{h,n+k}$ if detected signals $x_{h,n+k}$

are missing:

$$\hat{x}_{h,n+k} = \frac{1}{N+1}\sum_{i=0}^{N} x_{h,n-i} \tag{16}$$

2) Force sensing: A force sensor is mounted on the robot gripper, and calibration systems are utilized to obtain the relationship between robot reference coordinate $\{R\}$ and force sensor coordinate $\{F\}$. The transformation matrix ${}^{R}_{F}T$ denotes force sensor coordinate with respect to robot reference coordinate, matrix ${}^{E}_{F}T$ denotes force sensor coordinate with respect to robot's end-effector coordinate, and matrix ${}^{R}_{E}T$ denotes robot's end-effector coordinate with respect to robot reference coordinate, so ${}^{R}_{F}T$ can be transformed as follows:

$${}^{R}_{F}T = {}^{R}_{E}T \cdot {}^{E}_{F}T \tag{17}$$

Then we transform control input of human from force sensor coordinate $\{F\}$ to robot reference coordinate $\{R\}$:

$${}^{R}f_r = {}^{R}_{F}T \cdot {}^{F}f_r \tag{18}$$

where ${}^{F}f_r$ denotes the external force in force sensor coordinate $\{F\}$, and ${}^{R}f_r$ denotes the external force in robot reference coordinate $\{R\}$, respectively. Moving average filters similar to that in (16) are designed to obtain smooth force data, and limit breadth filter is utilized to deal with disturbances from external environment and mechanical friction.

B. Observer design and human motion intention estimation

For estimating control input of human f_h, we develop an observer in this section, which provides feasibility in scenarios where it is inconvenient to directly measure human force. We rewrite (2) in state-space form as follows:

$$\begin{aligned}
&\dot{\delta} = M_1\delta + N_1 f_h + N_1(f_r + G_o(x_o)) \\
&\delta = [x_o, \dot{x}_o]^T, M_1 = \begin{bmatrix} 0_n & 1_n \\ 0_n & -M_o(x_o)^{-1}C_o(x_o,\dot{x}_o) \end{bmatrix} \\
&N_1 = [0_n, M_o(x_o)^{-1}]
\end{aligned} \tag{19}$$

where 0_n denotes a matrix with all zero elements, and 1_n denotes an identity matrix. Object position x_0 can be calculated based on the relationship between human motion x_h and robot gripper position x_r as follows:

$$x_o = x_h + \frac{x_r - x_h}{2} \tag{20}$$

We design the following observer as:

$$\dot{\hat{\delta}} = M_1\hat{\delta} + N_1\hat{f}_h + N_1(f_r + G_o) - L(\hat{\delta} - \delta) \tag{21}$$

where $\hat{\delta}, \hat{f}_h$ denote the estimates of δ, f_h, respectively. And L denotes a positive definite matrix. We rewrite the human's control input (3) as follows:

$$f_h = -K_h(x_h - x_e) = -K_h x_h + K_h x_e = -K_h x_h + A_h \tag{22}$$

where we define $A_h = K_h x_e$ and neither human motion intention x_e nor human stiffness K_h is known for robot. From (22), we can obtain that

$$\hat{f}_h = -\hat{K}_h x_h + \hat{A}_h \tag{23}$$

where $\hat{A}$ and $\hat{K}_h$ denote the estimates of A and K_h, respectively. From (22) and (23), we obtain $\tilde{f}_h$ as follows:

$$\tilde{f}_h = -\tilde{K}_h x_h + \tilde{A}_h \tag{24}$$

where $\tilde{\bullet}$ denotes the estimation error of $\bullet$, i.e., $\tilde{\bullet} = \hat{\bullet} - \bullet$. From (21) and (19), we obtain the observation error system as follows:

$$\dot{\tilde{\delta}} = M_1\tilde{\delta} + N_1\tilde{f}_h - L\tilde{\delta} \tag{25}$$

And we design the following updating law for parameters in (23):

$$\begin{aligned}
\dot{\hat{A}}_h &= -N_1^T\tilde{\delta} + \beta\hat{f}_h - (\dot{x}_h - \dot{x}_d) \\
\dot{\hat{K}}_h &= (N_1^T\tilde{\delta} - \beta\hat{f}_h + (\dot{x}_h - \dot{x}_d))x_h^T
\end{aligned} \tag{26}$$

where β denotes a positive constant. When $\hat{f}_h$ is obtained, x_d can be calculated by the following updating law:

$$\dot{x}_d = \beta\hat{f}_h \tag{27}$$

Human motion intention x_e can be estimated based on control input of human f_h, and control input of human f_h has been estimated by our proposed observer, so x_e can be estimated. We set the robot's desired trajectory x_d in (8) based on the updating law (27), which means that robot is estimating the human motion intention online, i.e., x_d is generated towards x_e. For better illustrating control and observer design, a block diagram is given in Fig. 2. Although a human-robot co-carrying task is studied as a specific application, the proposed method can be extended to other scenarios of physical human-robot interaction that rely on force and visual sensing, such as human-robot collaborative assembly, tele-operation, sawing, etc.

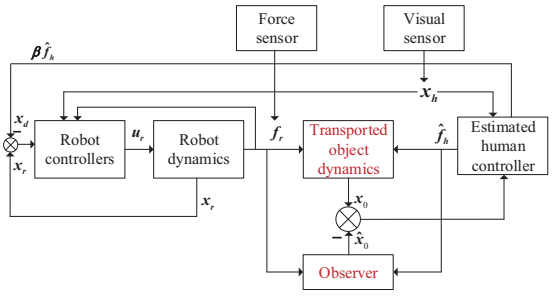


Fig. 2: The proposed control structure.

IV. Simulations

A. Simulation settings

We consider a scenario where human and robot perform co-transporting tasks in $X-Y$ plane. The object is chosen as a 0.1m long board which is located parallel to X-axis. In this task, rotation is not considered and only translational motion is involved. A human hand grasps one end of the board, and the other end of the board is held by the robot gripper. The

task objective is to move the object from an initial position to a target position.

We consider the robot as a simple two-link manipulator, where the length and mass of the first link are set as 0.3443m, 4.318kg, and the length and mass of the second link are set as 0.3443m, 2.152kg. By Lagrange equation, dynamics parameters $M_r(x_r)$, $C_r(x_r, \dot{x}_r)$ and $G_r(x_r)$ in (1) can be calculated based on the robot's physical parameters for simulating robot's position and velocity in the task space and the detailed expression can be found in [43]. We consider the initial position vector of robot gripper $x_r(0) = [0.2\text{m}, 0.25\text{m}]^{\text{T}}$ and the initial human hand position vector $x_h(0) = [0.1\text{m}, 0.25\text{m}]^{\text{T}}$. In (3), we consider human motion intention vector $x_e = [0.3\text{m}, 0.35\text{m}]^{\text{T}}$, and human arm stiffness matrix $K_h = [0.115\text{N/m}, 0; 0, 0.258\text{N/m}]$. We consider that the human arm arrives at x_e in 30s, which generates a prescribed human trajectory x_h in Fig. 3(a), and $x_h(1)$ and $x_h(2)$ denote uniform linear motions on X-axis and Y-axis. According to (3), the control input of human f_h can be simulated when x_h, K_h and x_e are available as shown in Fig. 3(b), and $C_o(x_o, \dot{x}_o)$ in (19) can be regarded as zero without rotation. The external force f_r can be obtained from (19), and $f_r(1)$ and $f_r(2)$ denote external forces on X-axis and Y-axis.

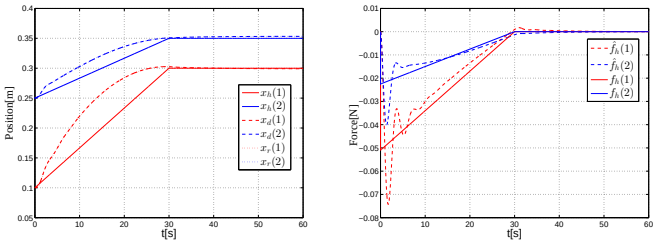
We set other crucial parameters as follows: the object mass in (19) is set as 0.6kg, and β in (27) is set as 0.3. In (6), RBFNN node number is set as 210, RBFNN centers are set in the region of $[-1, 1]$, and we define the initial value of the RBFNN weights θ_i as 0, positive definite gain matrices $\Gamma_1 = \Gamma_2 = 10I_{210\cdot 210}$, $\sigma_1 = 2.2$ and $\sigma_2 = 0.9$. In (8), the proportional gain matrix K_p is defined as $[10, 0; 0, 10]$, and the differential gain matrix K_d is defined as $[5, 0; 0, 5]$. We define L in (21) as follows:

$$L = 1.51 \cdot \begin{bmatrix} 0.103, 0.1, 0.1, 0.0 \\ 0.19, 0.6, 0.1, 0.2 \\ 0.08, 0.0, 0.2, 0.0 \\ 0.0, 0.2, 0.0, 0.7 \end{bmatrix} \tag{28}$$

B. Simulation results

Simulation results about robot's desired motion x_d, gripper position x_r and human motion x_h are shown in Fig. 3(a). It depicts that robot can estimate human motion intention accurately and generates a desired motion to perform the task successfully. We can conclude that the robot gripper position x_r tracks robot's desired motion x_d accurately under our proposed controller in (8), and tracking errors converge to zero on X-axis and Y-axis. As shown in Fig. 3(a), the motion synchronization of human and robot can be achieved. Indicated from Fig. 3(b), it is obvious that $\hat{f}_h$ can estimate f_h well which illustrates the effectiveness of our proposed observer (21).

For evaluating the robustness of our proposed method, we set three different human motion intention vectors as $x_{e1} = [0.3\text{m}, 0.35\text{m}]^{\text{T}}$, $x_{e2} = [0.3\text{m}, 0.35\text{m}]^{\text{T}}$ and $x_{e3} = [0.4\text{m}, 0.35\text{m}]^{\text{T}}$. Seen from Fig. 4, motion synchronization can be achieved when human motion intentions are different.



(a) Positions of x_h, x_d and x_r. (b) f_h and $\hat{f}_h$.

Fig. 3: Simulation results.

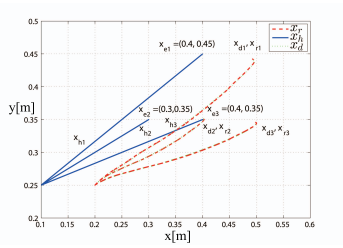


Fig. 4: Co-transporting considering different human motion intentions.

V. Experiments

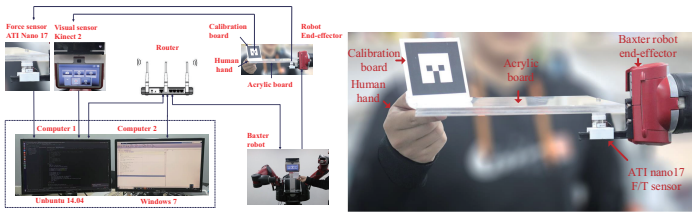
A. Experiment settings

As shown in Fig. 5(a), the right arm of Baxter robot is employed to cooperate with human to perform co-transporting tasks. Angles, angular velocities and torques can be obtained by sensors in all seven joints. Detailed introduction about Baxter robot can be referred to [44]. Considering both accuracy and computation efficiency, two computers are utilized in the experiments. One computer is used to calculate the feedforward input u_{rf} of NN compensation in (4) by Matlab Simulink, and transfer the compensation values to the other computer by UDP communications. The other computer is used to receive sensory information from Baxter robot, visual sensor and force sensor, calculate feedback control input u_{rb} and generate control input u_{rNN} to control the robot by Baxter Robot Operating System Software Development Kit (RSDK) in Ubuntu 14.04 LTS. Note that Jacobian matrix transpose J^T can be obtained from Python Kinematics and Dynamics Library (PyKDL), and control torque vector of seven joints is calculated by $J^T u_{rNN}$.

Kinect 2 3D depth camera is utilized as a visual sensor for obtaining human motion x_h. Kinect 2 contains a color camera, a depth sensor and four microphone arrays and provides capabilities in three dimensional (3D) motion capture and voice recognition. It is mounted on the hand of Baxter robot. Quick response detection method is applied in Kinect 2 to obtain the 3D position of the calibration board, which has been fixed on the human side of the board. We can obtain 3D locations of human hand x_h as the position of calibration board seen from Fig. 5(b). A calibrated force-torque (F/T) sensor ATI nano17 is used to obtain 6-DOF forces and torques on the robot gripper. Indicated from Fig. 5(b), robot gripper and human hand carry an acrylic board. Coordinate conversion

and filters are utilized in processing collected visual and force sensing information which have been described in Section III.

Then we utilize (15) to transform ${}^{K}x_h$ to ${}^{R}x_h$, and utilize (18) to transform ${}^{F}f_r$ to ${}^{R}f_r$. ${}^{R}x_h$ and ${}^{R}f_r$ are utilized for further controller design in robot reference coordinates.



(a) Experiment platform. (b) The co-transporting scenario.

Fig. 5: Experiment platform and co-transporting scenario.

B. Experiment results

1) Results with the proposed method: The experiment has been performed with 4 subjects with ages ranging from 20 to 32. The group is formed of 3 males and 1 female all right-handed, who are from our university, and they have robotic research experiences but are blind to experiment settings. As shown in Fig. 6, human subject B and Baxter robot move the board jointly along the direction indicated by red arrows, and synchronous motions in *up*, *right*, *down*, *left* and *diagonal* directions in order in Y-Z plane demonstrate the effectiveness of our proposed method. Indicated from Fig. 7(a), human subject A/B/C/D and robot co-transport the object in the *diagonal* direction from the same initial positions to the target positions. Due to human motion uncertainties in real applications, we only ensure similar x_h of four subjects. Fig. 7(b) depicts that the task can be accomplished by different human subjects. Sensor noises and rotations around X-axis result in non-smooth curves of x_h and x_r, which may influence experiment results and cannot be avoided. Fig. 7(b) shows that motion synchronization of robot and Subject B can be achieved on Y-axis and Z-axis (the board length is deducted from x_r on Y-axis for better comparison in figures). Mean squared error (MSE) is employed to evaluate the robustness of our proposed method, which is defined as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{k=1}^{n}[x_{r,k} - x_{h,k} - x_b]^2 \tag{29}$$

where x_b denotes the vector of the board dimension $[l, 0]^T$, l=0.24m denoting board length on Y-axis. We have added a criterion "Motion Smoothness (MS)" and define it as the distance between the upper and lower envelopes of the robot motion curve. The results of MSE and MS for different subjects are shown in Table I, which illustrates a small tracking error for all subjects and smooth movements. Fig. 8(a) shows that the external forces on robot are smooth and the co-carrying processes are stable. Limited by experimental equipment, f_h cannot be measured directly in our experiment, so (19) is utilized to approximatively calculate f_h on the end of the board. Table II shows that f_h of four subjects are small and continuous under our proposed method.

TABLE I: MSE and MS of 4 subjects.

MSE/MS \ subject	A	B	C	D
$\mathrm{MSE_Y(cm^2)}$	0.255	0.149	0.124	0.223
$\mathrm{MS_Y(cm)}$	0.047	0.024	0.016	0.015
$\mathrm{MSE_Z(cm^2)}$	1.644	1.035	0.529	0.498
$\mathrm{MS_Z(cm)}$	0.020	0.018	0.023	0.022

TABLE II: Average and maximal f_h of 4 subjects on Y, Z-axes.

force \ subject	A	B	C	D
Y_{average}(N)	0.806	0.791	0.964	0.774
Y_{max}(N)	4.512	2.603	5.349	2.516
Z_{average}(N)	1.332	1.520	1.715	1.268
Z_{max}(N)	3.779	3.805	4.021	2.599

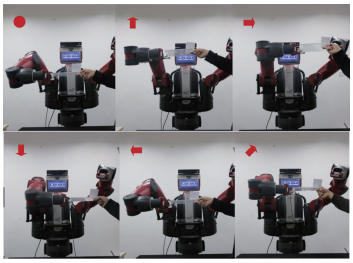
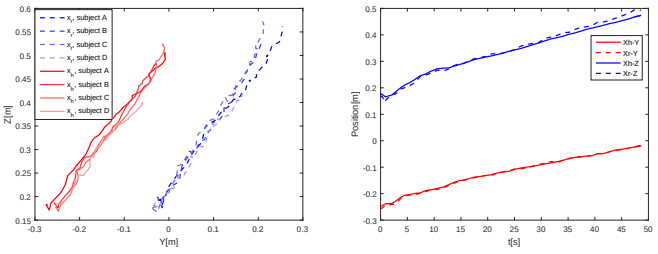
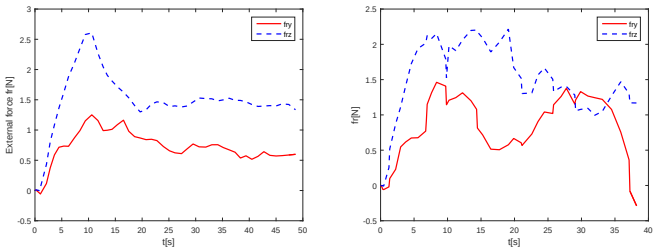


Fig. 6: Subject B performs a co-transporting task with Baxter robot along the direction of red arrows.



(a) Co-transporting considering different human subjects. (b) Motion synchronization of robot and Subject B.

Fig. 7: Human and robot motions.



(a) External force f_r using the proposed method. (b) External force f_r using PI.

Fig. 8: External force.

2) Comparisons: We compare our proposed method with VS (Visual Servoing method) and PI (Passive Impedance method) in this section. VS is utilized only with visual sensors to achieve motion tracking, and PI relies on sensory information from only force/torque senors to achieve compliant interactive behavior.

VS control method is redesigned according to (8) to make robot track x_d towards human motion x_h directly rather than human motion intention x_e, and a PD (proportional differential)-based controller $\tau_{\text{vision}} = -K_P(x_r - x_d) - K_D(\dot{x}_r - \dot{x}_d)$ is employed for tracking. The co-transporting task can not be performed successfully only by VS method, because there exist regid connections between human, board and robot. Therefore, we consider the tracking performance under VS without co-carrying the board for comparison. Figs. 9(a) and 9(b) show motion comparison (after 10s) under different methods (our proposed method, VS and PI), and the board length is deducted from x_r on Y-axis for better comparison in figures. Seen from Figs. 9(a) and 9(b), there exists a delay in robot motions when VS controller is involved.

A traditional PI control method is used for comparison, which is widely used in co-carrying tasks [45]. The desired impedance model of robot is designed as $f_r = D_d\dot{x}_r + K_d(x_r - x_d)$, where D_d and K_d denote damping and stiffness matrices, and we design K_d as zero for achieving compliant behaviors in experiments. Seen from Figs. 9(a) and 9(b), the robot motion under PI is less smooth and even includes oscillations. Indicated from Fig. 8(b), external force f_r is less smooth than that under our proposed method shown in Fig. 8(a), and human subjects report that they found the interaction uncomfortable. We conclude from Table III that f_h under our proposed method is smaller than that under PI, which illustrates that human subjects cost less efforts in the task. From Table IV, we find that our proposed method shows better collaborative performance compared with VS and PI in the task. Notably, Table IV shows better tracking results on Y-axis using VS, but there exists a delay in co-carrying tasks shown in Figs. 9(a) and 9(b), which may lead human to cost more effort for human in co-carrying tasks.

TABLE III: Average and maximal f_h on Y, Z-axes.

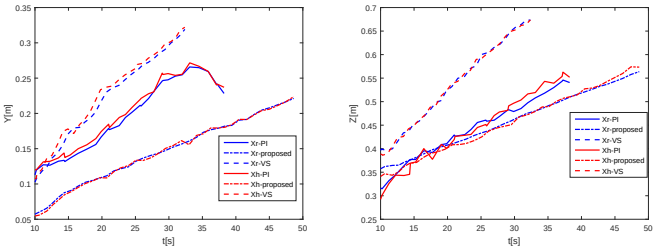
force \ method	PI	proposed
Y_{average}(N)	1.054	0.791
$Y_{\max}$(N)	3.177	2.603
Z_{average}(N)	1.597	1.520
$Z_{\max}$(N)	7.542	3.805

VI. Conclusions

A hybrid framework using visual and force sensing in human-robot co-transporting tasks has been proposed in our paper. Visual sensing has been employed to obtain human motion and force sensing has been used to measure external forces on robots. An observer has been designed for estimating control input of human, and robot's desired motion has been designed based on the observer towards human motion intention. An adaptive controller has been proposed for improving tracking accuracy, and online NNs have been used to compensate for uncertainties in robot's dynamics. The proposed framework has been validated by comparative simulations and experimental co-carrying tasks.

TABLE IV: MSE and MS under our proposed method compared with PI and VS.

MSE/MS \ subject	proposed	PI	VS
$\text{MSE}_\text{Y}(\text{cm}^2)$	0.149	3.751	0.734
$\text{MS}_\text{Y}(\text{cm})$	0.026	0.040	0.027
$\text{MSE}_\text{Z}(\text{cm}^2)$	1.035	1.813	1.017
$\text{MS}_\text{Z}(\text{cm})$	0.025	0.029	0.036



(a) Motion comparison on Y-axis. (b) Motion comparison on Z-axis.

Fig. 9: Motion comparison.

VII. Limitations and Future Works

In future works, co-transporting tasks such as human-robot collaborative assembly, tele-operation, sawing, etc will be further designed to evaluate our proposed framework. The calibration board was used for obtaining human motion and it was sometimes not recognized in the experiments. Therefore, more robust machine learning methods [46] [47] [48] will be investigated to localize the human hand by visual sensors without the calibration board. For instance, the Faster-RCNN detection algorithm can be used to localize the human hand position, while SiamRPN tracking algorithm can be utilized to realize real-time tracking. In this work, the interactive experience was described by human subjects verbally and is evaluated by human control efforts. In future works, force sensors will be mounted on the human side of the board for evaluating human interactive experience using objective measures. Human user studies with questionnaires and subjective measures will be also designed.

Appendix A

We consider Lyapunov function candidates V including V_e, V_k and V_x as follows:

$$V = V_e + V_k + V_x$$

$$V_e = \frac{1}{2}(\dot{e}^T M_r(x_r)\dot{e} + e^T K_p e) + \frac{1}{2}\sum_{i=1}^{n} \tilde{\theta}_i^{\ T} \Gamma_i^{-1} \tilde{\theta}_i$$
$$V_k = \frac{1}{2}(\dot{\tilde{x}}_o^T \dot{\tilde{x}}_o + \tilde{f}_h^T \tilde{f}_h + (\mathrm{vec}(\tilde{K}_h)^T \mathrm{vec}(\tilde{K}_h))$$
$$V_x = \frac{1}{2}(x_h - x_e)^T K_h (x_h - x_e) \tag{30}$$

Differentiating V_e yields:

$$\dot{V}_e = \dot{e}^T M_r(x_r)\ddot{e} + \frac{1}{2}\dot{e}^T \dot{M}_r(x_r)\dot{e} + \dot{e}^T K_p e + \sum_{i=1}^{n} \tilde{\theta}_i^T \Gamma_i^{-1} \dot{\tilde{\theta}}_i \tag{31}$$

where $\frac{1}{2}(M_r(x_r) - 2C_r(x_r, \dot{x}_r))$ is a skew-symmetric matrix [43], so we can obtain that $\frac{1}{2}e^T(M_r(x_r) - 2C_r(x_r, \dot{x}_r))e = 0$. Substituting (9) and (6), we rewrite (31) according to *Young's inequality* as follows:

$$\begin{aligned}\dot{V}_e =& \dot{e}^T(M_r(x_r)\ddot{e} + C_r(x_r, \dot{x}_r)\dot{e} + K_p e) \\ &+ \sum_{i=1}^{n} \tilde{\theta}_i^T \Gamma_i^{-1}\{-\Gamma_i[S_i(Z_i)\dot{e}_i + \delta_i \hat{\theta}_i]\} \\ &\leq -\dot{e}^T K_d \dot{e} + (\dot{x}_h - \dot{x}_d)^T \hat{f}_h + \sum_{i=1}^{n} \frac{\sigma_i}{2}(||\theta_i^*||^2 - ||\tilde{\theta}_i||^2)\end{aligned} \tag{32}$$

where $K_q \geq ||\varepsilon(Z)||$, then differentiating V_k yields:

$$\begin{aligned}\dot{V}_k =& \tilde{\delta}^T \dot{\tilde{\delta}} + \tilde{A}_h^T \dot{\hat{A}}_h + \mathrm{vec}(\tilde{K}_h)^T \mathrm{vec}(\dot{\tilde{K}}_h) \\ =& \tilde{\delta}^T N_1 \tilde{f}_h - \tilde{\delta}^T (L - M_1)\tilde{\delta} - \tilde{A}_h^T N_1^T \tilde{\delta} + \tilde{A}_h^T \beta f_h - \dot{e}^T \tilde{f}_h \\ &+ \mathrm{vec}(\tilde{K}_h)^T \mathrm{vec}(N_1^T \tilde{\delta} x_h^T) - \mathrm{vec}(\tilde{K}_h)^T \mathrm{vec}(\beta f_h x_h^T) \\ =& -\tilde{\delta}^T (L - M_1)\tilde{\delta} + \beta \tilde{f}_h^T f_h - (\dot{x}_h - \dot{x}_d)^T \tilde{f}_h\end{aligned} \tag{33}$$

Differentiating V_x yields:

$$\dot{V}_x = \dot{x}_h^T K_h (x_h - x_e) = -\dot{x}_h^T f_h \tag{34}$$

Adding $\dot{V}_x$ to $\dot{V}_e$ we obtain:

$$\begin{aligned}\dot{V}_e + \dot{V}_x \leq& -\dot{e}^T K_d \dot{e} + (\dot{x}_h - \dot{x}_d)^T f_h + (\dot{x}_h - \dot{x}_d)^T \tilde{f}_h - \dot{x}_h^T f_h \\ &+ F \\ \leq& -\dot{e}^T K_d \dot{e} + (\dot{x}_e^T - \dot{x}_d^T) f_h + \dot{e}^T \tilde{f}_h - \dot{x}_h^T f_h \\ &+ (\dot{x}_h - \dot{x}_d)^T \tilde{f}_h + F \\ \leq& -\dot{e}^T K_d \dot{e} - \beta f_h^T f_h - \beta \tilde{f}_h^T f_h + (\dot{x}_h - \dot{x}_d)^T \tilde{f}_h + F\end{aligned} \tag{35}$$

where $F = \sum_{i=1}^{n} \frac{\sigma_i}{2}(||\theta_i^*||^2 - ||\tilde{\theta}_i||^2)$. So we obtain $\dot{V}$ as follows:

$$\begin{aligned}\dot{V} =& \dot{V}_e + \dot{V}_k + \dot{V}_x \\ \leq& -\dot{e}^T K_d \dot{e} - \beta f_h^T f_h - \beta \tilde{f}_h^T f_h - \tilde{\delta}^T (L - M_1)\tilde{\delta} + \beta \tilde{f}_h^T f_h \\ &+ \sum_{i=1}^{n} \frac{\sigma_i}{2}(||\theta_i^*||^2 - ||\tilde{\theta}_i||^2) \\ \leq& -\dot{e}^T K_d \dot{e} - \beta f_h^T f_h - \tilde{\delta}^T (L - M_1)\tilde{\delta} - \sum_{i=1}^{n} \frac{\sigma_i}{2}||\tilde{\theta}_i||^2 \\ &+ \sum_{i=1}^{n} \frac{\sigma_i}{2}||\theta_i^*||^2\end{aligned} \tag{36}$$

Then we can conclude that variables $\dot{e}$, f_h and $\tilde{\delta}$ are bounded and satisfy a condition as follows

$$\begin{aligned}&\lambda_{K_d}||\dot{e}||^2 + \beta_{\min}||f_h||^2 + \lambda_{L-M_1}||\tilde{\delta}||^2 + \frac{\sigma_i}{2}||\mathrm{vec}(\tilde{\theta}_i)||^2 \\ &\leq \frac{\sigma_i}{2}||\mathrm{vec}(\theta_i^*)||^2\end{aligned} \tag{37}$$

where λ_{K_d} and λ_{L-M_1} are the minimal eigenvalues of K_D and K_{L-M_1}, respectively, $\beta_{\min}$ denotes the minimal value of β, and vec$(\cdot)$ stands for the column vectorization operation. It follows that $\dot{e}$, f_h and $\tilde{x}_o$ can be made arbitrarily small by choosing sufficiently large λ_{K_d}, λ_L and $\beta_{\min}$. If θ_i^* is zero, we can conclude that $\dot{V} = 0$ when $\dot{e} = 0$, $f_h = 0$ and $\tilde{\delta} = 0$. We consider $\dot{e} = 0$ in robotic dynamics (9), and obtain that $K_P e = 0$, so $e = 0$ and $x_r = x_d$. Indicated from $f_h = 0$, we can obtain $x_h = x_e (K_h \neq 0)$ or $K_h = 0$. By considering $\tilde{\delta} = 0$ in (25) we can obtain that $\tilde{f}_h = 0$ which means that control input of human can be obtained.

The above inequality (37) can be proved by contradiction: assuming the above inequality is invalid yields $\dot{V} < 0$ and thus V decreases iteratively. This indicates that $||\dot{e}||$, $||f_h||$, $||\tilde{\delta}||$ and $||\mathrm{vec}(\tilde{\theta}_i)||$ (and thus the left-hand side of the above inequality) become even smaller, which contradicts the hypothesis.

REFERENCES

[1] B. Siciliano and O. Khatib, *Springer handbook of robotics*. Springer, 2016.

[2] Z. Li, B. Huang, A. Ajoudani, C. Yang, C.-Y. Su, and A. Bicchi, "Asymmetric bimanual control of dual-arm exoskeletons for human-cooperative manipulations," *IEEE Transactions on Robotics*, vol. 34, no. 1, pp. 264–271, 2017.

[3] M. Khoramshahi and A. Billard, "A dynamical system approach to task-adaptation in physical human–robot interaction," *Autonomous Robots*, pp. 1–20, 2018.

[4] Y. Li and S. S. Ge, "Human–robot collaboration based on motion intention estimation," *IEEE/ASME Transactions on Mechatronics*, vol. 19, no. 3, pp. 1007–1014, 2014.

[5] M. S. Erden and T. Tomiyama, "Human-intent detection and physically interactive control of a robot without force sensors," *IEEE Transactions on Robotics*, vol. 26, no. 2, pp. 370–382, 2010.

[6] L. Peternel, N. Tsagarakis, and A. Ajoudani, "Towards multi-modal intention interfaces for human-robot co-manipulation," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, pp. 2663–2669, IEEE, 2016.

[7] T. Lenzi, S. M. M. De Rossi, N. Vitiello, and M. C. Carrozza, "Intention-based emg control for powered exoskeletons," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2180–2190, 2012.

[8] D. P. Losey and M. K. OMalley, "Trajectory deformations from physical human–robot interaction," *IEEE Transactions on Robotics*, vol. 34, no. 1, pp. 126–138, 2018.

[9] C. Yang, C. Zeng, P. Liang, Z. Li, R. Li, and C.-Y. Su, "Interface design of a physical human-robot interaction system for human impedance adaptive skill transfer," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 1, pp. 329–340, 2018.

[10] H. Li, S. Zhao, W. He, and R. Lu, "Adaptive finite-time tracking control of full state constrained nonlinear systems with dead-zone," *Automatica*, vol. 100, pp. 99–107, 2019.

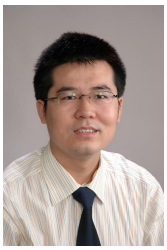
[11] W. He, W. Ge, Y. Li, Y.-J. Liu, C. Yang, and C. Sun, "Model identification and control design for a humanoid robot," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 1, pp. 45–57, 2017.

[12] J. Chen and H. Qiao, "Muscle-synergies-based neuromuscular control for motion learning and generalization of a musculoskeletal system," *IEEE Transactions on Systems, Man, and Cybernetics: Systems, in press, DOI: 10.1109/TSMC.2020.2966818*, 2020.

[13] H. Qiao, M. Wang, J. Su, S. Jia, and R. Li, "The concept of attractive region in environment and its application in high-precision tasks with low-precision systems," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 5, pp. 2311–2327, 2014.

[14] G. Xie, A. Shangguan, R. Fei, W. Ji, W. Ma, and X. Hei, "Motion trajectory prediction based on CNN-LSTM sequential model," *SCIENCE CHINA Information Sciences, in press, DOI: 10.1007/s11432-019-2761-y*, 2020.
[15] D. Huang, Y. Fu, N. Qin, and S. Gao, "Fault diagnosis of high-speed train bogie based on LSTM neural network," *SCIENCE CHINA Information Sciences*, vol. 64, no. 1, p. 119203, 2020.
[16] M. Bortolini, M. Faccio, F. G. Galizia, M. Gamberi, and F. Pilati, "Design, engineering and testing of an innovative adaptive automation assembly system," *Assembly Automation*, vol. 40, no. 3, pp. 531–540, 2020.
[17] G. Wu, J. Sun, and J. Chen, "Optimal linear quadratic regulator of switched systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2898–2904, 2019.
[18] H. Li, Y. Gao, P. Shi, and H.-K. Lam, "Observer-based fault detection for nonlinear systems with sensor fault and limited communication capacity," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2745–2751, 2016.
[19] B. Gao, X. Li, W. L. Woo, and G. yun Tian, "Physics-based image segmentation using first order statistical properties and genetic algorithm for inductive thermography imaging," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2160–2175, 2018.
[20] X. Chen, Y. Feng, and C.-Y. Su, "Adaptive control for continuous-time systems with actuator and sensor hysteresis," *Automatica*, vol. 64, pp. 196–207, 2016.
[21] X. Chen, C.-Y. Su, Z. Li, and F. Yang, "Design of implementable adaptive control for micro/nano positioning system driven by piezoelectric actuator," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 10, pp. 6471–6481, 2016.
[22] X. Yu, S. Zhang, L. Sun, Y. Wang, C. Xue, and B. Li, "Cooperative control of dual-arm robots in different human-robot collaborative tasks," *Assembly Automation*, vol. 40, no. 1, pp. 95–104, 2019.
[23] C. Yang, G. Ganesh, S. Haddadin, S. Parusel, A. Albu-Schaeffer, and E. Burdet, "Human-like adaptation of force and impedance in stable and unstable interactions," *IEEE Transactions on Robotics*, vol. 27, no. 5, pp. 918–930, 2011.
[24] B. Xu, "Composite learning control of flexible-link manipulator using NN and DOB," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 11, pp. 1979–1985, 2017.
[25] J. Na, Y. Li, Y. Huang, G. Gao, and Q. Chen, "Output feedback control of uncertain hydraulic servo systems," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 1, pp. 490–500, 2019.
[26] K. Zhang and Y. Shi, "Adaptive model predictive control for a class of constrained linear systems with parametric uncertainties," *Automatica*, vol. 117, p. 108974, 2020.
[27] J. Na, Y. Huang, X. Wu, G. Gao, G. Herrmann, and J. Z. Jiang, "Active adaptive estimation and control for vehicle suspensions with prescribed performance," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 6, pp. 2063–2077, 2017.
[28] H. Lin, B. Zhao, D. Liu, and C. Alippi, "Data-based fault tolerant control for affine nonlinear systems through particle swarm optimized neural networks," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 4, pp. 954–964, 2020.
[29] Y. Ren, M. Chen, and J. Liu, "Bilateral coordinate boundary adaptive control for a helicopter lifting system with backlash-like hysteresis," *Science China Information Sciences*, vol. 63, no. 119203, 2020.
[30] Y. Zhang, J. Sun, H. Liang, and H. Li, "Event-triggered adaptive tracking control for multiagent systems with unknown disturbances," *IEEE Transactions on Cybernetics, DOI: 10.1109/TCYB.2018.2869084*, 2018.
[31] Y. Li, K. P. Tee, R. Yan, W. L. Chan, and Y. Wu, "A framework of human–robot coordination based on game theory and policy iteration," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1408–1418, 2016.
[32] A. Bussy, P. Gergondet, A. Kheddar, F. Keith, and A. Crosnier, "Proactive behavior of a humanoid robot in a haptic transportation task with a human partner," in *RO-MAN, 2012 IEEE*, pp. 962–967, 2012.
[33] L. Rozo, S. Calinon, D. Caldwell, P. Jimenez, and C. Torras, "Learning collaborative impedance-based robot behaviors," in *Twenty-Seventh AAAI Conference on Artificial Intelligence*, pp. 1422–1428, 2013.
[34] L. Rozo, S. Calinon, D. G. Caldwell, P. Jimenez, and C. Torras, "Learning physical collaborative robot behaviors from human demonstrations," *IEEE Transactions on Robotics*, vol. 32, no. 3, pp. 513–527, 2016.
[35] D. Cehajic, S. Erhart, and S. Hirche, "Grasp pose estimation in human-robot manipulation tasks using wearable motion sensors," in *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1031–1036, 2015.
[36] D. Ćehajić, S. Hirche, *et al.*, "Estimating unknown object dynamics in human-robot manipulation tasks," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pp. 1730–1737, IEEE, 2017.
[37] D. J. Agravante, A. Cherubini, A. Bussy, P. Gergondet, and A. Kheddar, "Collaborative human-humanoid carrying using vision and haptic sensing," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pp. 607–612, IEEE, 2014.
[38] L. Yang, Y. Li, and D. Huang, "Motion synchronization in human-robot co-transport without force sensing," in *2018 37th Chinese Control Conference (CCC)*, pp. 5369–5374, IEEE, 2018.
[39] Y. Chen, W. Wang, Z. Abdollahi, Z. Wang, J. Schulte, V. Krovi, and Y. Jia, "A robotic lift assister: a smart companion for heavy payload transport and manipulation in automotive assembly," *IEEE Robotics & Automation Magazine*, vol. 25, no. 2, pp. 107–119, 2018.
[40] W. Wang, R. Li, Y. Chen, Z. M. Diekel, and Y. Jia, "Facilitating human–robot collaborative tasks by teaching-learning-collaboration from human demonstrations," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 2, pp. 640–653, 2018.
[41] Z. Zhang, W. Wang, Y. Chen, Y. Jia, and G. Peng, "Prediction of human actions in assembly process by a spatial-temporal end-to-end learning model," tech. rep., SAE Technical Paper, 2019-01-0509, 2019.
[42] N. Andreff, R. Horaud, and B. Espiau, "Robot hand-eye calibration using structure-from-motion," *The International Journal of Robotics Research*, vol. 20, no. 3, pp. 228–248, 2001.
[43] W. He, S. S. Ge, Y. Li, E. Chew, and Y. S. Ng, "Neural network control of a rehabilitation robot by state and output feedback," *Journal of Intelligent & Robotic Systems*, vol. 80, no. 1, pp. 15–31, 2015.
[44] C. Yang, Y. Jiang, Z. Li, W. He, and C.-Y. Su, "Neural control of bimanual robots with guaranteed global stability and motion precision," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1162–1171, 2016.
[45] A. Mörtl, M. Lawitzky, A. Kucukyilmaz, M. Sezgin, C. Basdogan, and S. Hirche, "The role of roles: Physical cooperation between humans and robots," *The International Journal of Robotics Research*, vol. 31, no. 13, pp. 1656–1674, 2012.
[46] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299, 2017.
[47] R. Alp Güler, N. Neverova, and I. Kokkinos, "Densepose: Dense human pose estimation in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7297–7306, 2018.
[48] H. Joo, T. Simon, and Y. Sheikh, "Total capture: A 3d deformation model for tracking faces, hands, and bodies," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8320–8329, 2018.

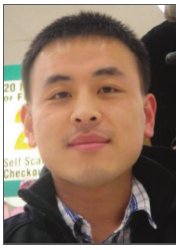
Xinbo Yu (S'16-M'20) received the B.E. degree in control technology and instrument from the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China, in 2013 and the Ph.D. degree in control science and engineering from the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China, in 2020. He is currently working as an associate professor in the Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing, China. His current research interests include adaptive neural networks control, robotics and human-robot interaction.



Wei He (S'09-M'12-SM'16) received his B.Eng. in automation and his M.Eng. degrees in control science and engineering from College of Automation Science and Engineering, South China University of Technology (SCUT), China, in 2006 and 2008, respectively, and his Ph.D. degree in control science and engineering from from Department of Electrical & Computer Engineering, the National University of Singapore (NUS), Singapore, in 2011.

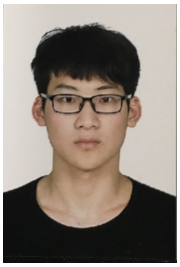
He is currently working as a full professor in School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China. He has co-authored 2 books published in Springer and published over 100 international journal and conference papers. He was awarded a Newton Advanced Fellowship from the Royal Society, UK in 2017. He was a recipient of the IEEE SMC Society Andrew P. Sage Best Transactions Paper Award in 2017. He is serving the Chair of IEEE SMC Society Beijing Capital Region Chapter. He is serving as an Associate Editor of *IEEE Transactions on Robotics*, *IEEE Transactions on Neural Networks and Learning Systems*, *IEEE Transactions on Control Systems Technology*, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *SCIENCE CHINA Information Sciences*, *IEEE/CAA Journal of Automatica Sinica*, *Neurocomputing* and an Editor of *Journal of Intelligent & Robotic Systems*. His current research interests include robotics, distributed parameter systems and intelligent control systems.

Qing Li received his B.S. degree from North China University of Science and Technology, Tangshan, China, in 1993 and the Ph.D degree in control theory and its applications from University of Science and Technology Beijing, Beijing, China, in 2000. He is currently a Professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China. He has been a visiting scholar at Ryerson University, Toronto, Canada, from February 2006 to February 2007. His research interests include intelligent control and intelligent optimization.



Yanan Li (S'10-M'14) is a Lecturer in Control Engineering with the Department of Engineering and Design, University of Sussex, UK. He received the B.Eng. and M.Eng. degrees from the Harbin Institute of Technology, China, in 2006 and 2008, respectively, and the PhD degree from the National University of Singapore, in 2013.

From 2015 to 2017, he has been a Research Associate with the Department of Bioengineering, Imperial College London, UK. From 2013 to 2015, he has been a Research Scientist with the Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore. Dr Li has active research in human-robot interaction and robot control and their applications in semi-autonomous industrial robots, tele-operation robots and rehabilitation robots. He has served as Technical Committee on Bio-mechatronics and Bio-robotics Systems and Technical Committee on Autonomous Bionic Robotic Aircraft, IEEE Systems, Man, and Cybernetics Society, and International Program Committee Member and Session Chair for several conferences in robotics and control. He has led and participated in several research projects funded by Xiamen Municipal Government, EU, Singapore SERC, MDA, etc.



Bin Li received his B.Eng. degree in automation from School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China, in 2019. He is currently pursuing the M.Eng. degree in control science and engineering from School of Automation and Electrical Engineering, University of Science and Technology Beijing. His current research interests include intelligent robot control, human-robot collaboration.